

# Optimal Sign Test for High Dimensional Location Parameters

Long Feng

*Northeast Normal University*

## Abstract

This article concerns tests for location parameters in cases where the data dimension is larger than the sample size. We propose a family of tests based on the optimality arguments in Le Cam (1986) under elliptical symmetric. The asymptotic normality of these tests are established. By maximizing the asymptotic power function, we propose an uniformly optimal test for all elliptical symmetric distributions. The optimality is also confirmed by a Monte Carlo investigation.

**Keywords:** High-dimensional data; Spatial sign; Uniformly optimal.

## 1 Introduction

Testing the population mean vector is a fundamental problem in statistics. A classical method to deal with this problem is the famous Hotelling's  $T^2$  test. However, it can not work in high dimensional settings because the sample covariance matrix is not invertible. With the rapid development of technology, various types of high-dimensional data have been generated in many areas, such as internet portals, microarray analysis. By replacing the Mahalanobis distance by the Euclidean distance, many modified Hotelling's  $T^2$  tests for high dimensional data are proposed in many literatures, such as Bai and Saranadasa (1996), Chen and Qin (2010), Srivastava (2009), Feng, et al. (2015b). However, the statistical performance of the

moment-based tests mentioned above would be degraded when the non-normality is severe, especially for heavy-tailed distributions.

Many nonparametric methods have been developed, as a reaction to the Gaussian approach of Hotelling's test, with the objective of extending to the multivariate context the classical univariate rank and signed-rank techniques. There are three main groups. One relies on componentwise rankings (Puri and Sen, 1971), but is not affine invariant. The second group is based on spatial signs and ranks with the so called Oja median (Oja, 2010). Some efforts have been devoted to extending this type of method to the high dimensional data. Wang, et al. (2015) propose a high dimensional spatial sign test by replacing the scatter matrix with identity matrix. Feng, et al. (2015a) also propose a scalar-invariant high dimensional sign test for the two sample location problem. They demonstrate that the multivariate sign and rank are still very efficient methods in constructing robust test in high dimension settings. The last group use the concept of interdirections (Randles, 1992). In an important work, Hallin and Paindaveine (2002) propose a class of tests based on interdirections and pseudo-Mahalanobis ranks. Depending on the score function considered, they allow for locally asymptotically maximin test at selected densities. However, to the best of our knowledge, there are no optimal tests for high dimensional location parameters.

In this article, we propose an uniformly optimal test for high dimensional data. Based on the optimality arguments in Le Cam (1986), we introduce a high dimensional form of the locally and asymptotically optimal testing procedure. The asymptotic normality of this class of tests are established. By maximizing the power function of these tests, we propose an uniformly optimal test for high dimensional location problem. In the multivariate case, the optimal score function deeply depends on the underlying distributions. However, the optimal weighted function for our high dimensional test is unique. So our proposed test procedure is uniformly optimal for the elliptical symmetric distributions. We also derive the asymptotic relative efficiency of our test with respect to Chen and Qin (2010)'s test and Wang, et al. (2015)'s test. It is not surprised that they are all no less than one for the elliptical symmetric distributions. And for the heavy tailed distributions, such as multivariate  $t$ -distributions or

mixture multivariate normal distributions, our test would perform eventually better than these two tests. Simulation studies also demonstrate these results.

## 2 Uniformly Optimal Test

### 2.1 High Dimensional Weighted Sign tests

Assume  $\{\mathbf{X}_i\}_{i=1}^n$  are i.i.d. random sample from  $p$ -variate elliptically symmetric distribution with density function  $\det(\Sigma)^{-1/2}g(\|\Sigma^{-1/2}(\mathbf{x} - \theta)\|)$  where  $\theta$  is the symmetry centers and  $\Sigma$  is the positive definite symmetric  $p \times p$  scatter matrices. We consider the following one sample testing problem

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta \neq 0. \quad (1)$$

When the dimension  $p$  is fixed, according to the local asymptotic normality theory (Le Cam, 1986), the form of locally and asymptotically optimal testing procedures for (1) under specified  $\Sigma$  and  $g$  is

$$Q_n = \frac{p}{nc_{p,g}} \sum_{i=1}^n \sum_{j=1}^n \psi_g(\|\Sigma^{-1/2} \mathbf{X}_i\|) \psi_g(\|\Sigma^{-1/2} \mathbf{X}_j\|) U(\Sigma^{-1/2} \mathbf{X}_i)^T U(\Sigma^{-1/2} \mathbf{X}_j),$$

where  $U(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|I(\mathbf{x} \neq \mathbf{0})$ ,  $\psi_g = -g'/g$  and  $c_{p,g}$  is a scaled parameter. Hallin and Paindaveine (2002) proposed a class of tests based on interdirections and pseudo-Mahalanobis ranks which are of the asymptotic form

$$R_n = \frac{2}{n(n-1)} \sum_{i < j} K(\|\Sigma^{-1/2} \mathbf{X}_i\|) K(\|\Sigma^{-1/2} \mathbf{X}_j\|) U(\Sigma^{-1/2} \mathbf{X}_i)^T U(\Sigma^{-1/2} \mathbf{X}_j),$$

$K(\cdot)$  is a continuous weighted function. However, the scatter matrix  $\Sigma$  is not available in high dimensional settings. Motivated by Bai and Saranadasa (1996) and Chen and Qin (2010), we simply replace  $\Sigma$  by  $\mathbf{I}_p$  and exclude the same term in  $R_n$ . We propose the following generally weighted sign test statistic:

$$W_n = \frac{2}{n(n-1)} \sum_{i < j} K(r_i) K(r_j) U(\mathbf{X}_i)^T U(\mathbf{X}_j),$$

where  $r_i = \|\mathbf{X}_i\|$ . Let  $K(t) = t$ ,  $W_n$  would be the one-sample high dimensional  $t$ -test statistic proposed in Chen and Qin (2010). Similarly, we can obtain the high dimensional sign test (Wang, et al. , 2015) with  $K(t) = 1$ . We will determine the optimal weighted function  $K(t)$  in the next section. First, we propose an asymptotic analysis for  $W_n$ .

Recently, there are many high dimensional scalar-invariant tests in literature (Park and Ayyala , 2013; Srivastava , 2009; Feng, et al. , 2015a,b). The idea is replacing  $\Sigma$  by its diagonal matrix. And then all the variables have the same scale. Here we also standardize each variables first by the estimated diagonal matrix in Feng, et al. (2015a), which make  $W_n$  invariant under the scale transformation. Details about the scalar-invariant test are given in the appendix. To expedite our discussion, we assume the diagonal matrix of  $\Sigma$  are known and equal to one without loss of generality.

The following conditions are needed.

$$(C1) \quad \text{tr}(\Sigma^4) = o(\text{tr}^2(\Sigma^2)) \text{ and } \text{tr}(\Sigma^2) - p = o(n^{-1}p^2).$$

$$(C2) \quad \nu_4 = O(\nu_2^2) \text{ where } \nu_l = E(K^l(r_i)).$$

The first condition in (C1) is similar to condition (3.8) in Chen and Qin (2010). Obviously, (C1) will hold if all the eigenvalues of  $\Sigma$  are bounded. The second condition in Condition (C1) is used to reduce the difference between the module  $\|\epsilon\|$  and  $\|\Sigma^{1/2}\epsilon\|$ . Then, we can get an explicit relationship between the variance of  $W_n$  and  $\Sigma$ . Condition (C2) is similar to Assumption 1 in Zou et al. (2014) if we choose  $K(t) = t^{-1}$ .

**Theorem 1** *Under Conditions (C1)-(C2) and  $H_0$ , as  $(p, n) \rightarrow \infty$ ,*

$$W_n/\sigma_n \xrightarrow{d} N(0, 1)$$

where  $\sigma_n^2 = 2n^{-2}p^{-2}\nu_2^2\text{tr}(\Sigma^2)$ .

Similar to Wang, et al. (2015), we propose the following ratio-consistent estimator of  $\sigma_n^2$

$$\hat{\sigma}_n^2 = 2n^{-4} \sum_{i \neq j} K^2(r_i) K^2(r_j) \{U(\mathbf{X}_i) - \boldsymbol{\mu}_{i,j}\}^T U(\mathbf{X}_j) \{U(\mathbf{X}_j) - \boldsymbol{\mu}_{i,j}\}^T U(\mathbf{X}_i),$$

where  $\boldsymbol{\mu}_{i,j} = \frac{1}{n-2} \sum_{k \neq i,j} U(\mathbf{X}_k)$ . And then we reject the null hypothesis if  $W_n/\hat{\sigma}_n > z_\alpha$  where  $z_\alpha$  is the upper  $\alpha$  quantile of  $N(0, 1)$ .

Next, we consider the asymptotic distribution of  $W_n$  under the alternative hypothesis

$$(C3) \quad \boldsymbol{\theta}^T \boldsymbol{\theta} = O(c_0^{-2} \sigma_n), \quad \boldsymbol{\theta}^T \boldsymbol{\Sigma} \boldsymbol{\theta} = o(np c_0^{-2} \sigma_n) \quad \text{where } c_0 = E\{K(r_i) r_i^{-1}\}.$$

Condition (C3) require the difference between  $\boldsymbol{\mu}$  and 0 is not large so that the variance of  $W_n$  is still asymptotic  $\sigma_n^2$ . It can be viewed as a high-dimensional version of the local alternative hypotheses.

**Theorem 2** *Under Conditions (C1)-(C3), as  $(p, n) \rightarrow \infty$ , we have*

$$\frac{W_n - c_0^2 \boldsymbol{\theta}^T \boldsymbol{\theta}}{\sigma_n} \xrightarrow{d} N(0, 1).$$

## 2.2 High Dimensional Optimal Sign test

According to Theorem 1 and 2, the asymptotic power of our weighted sign test becomes

$$\beta_{\text{ws}}(||\boldsymbol{\theta}||) = \Phi \left( -z_\alpha + \frac{[E\{K(r_i) r_i^{-1}\}]^2}{E\{K^2(r_i)\}} \frac{pn \boldsymbol{\theta}^T \boldsymbol{\theta}}{\sqrt{2\text{tr}(\boldsymbol{\Sigma}^2)}} \right).$$

The power function of  $W_n$  is an increasing function of  $\frac{[E\{K(r_i) r_i^{-1}\}]^2}{E\{K^2(r_i)\}}$ . By the Cauchy inequality, we have

$$\frac{[E\{K(r_i) r_i^{-1}\}]^2}{E\{K^2(r_i)\}} \leq \frac{E\{K^2(r_i)\} E(r_i^{-2})}{E\{K^2(r_i)\}} = E(r_i^{-2}).$$

The maximum of  $\beta_{\text{ws}}(||\boldsymbol{\theta}||)$  is  $E(r_i^{-2})$  with maximizer  $K(t) = t^{-1}$ . Consequently, we propose the following high dimensional optimal sign test

$$O_n = \frac{2}{n(n-1)} \sum_{i < j} r_i^{-1} r_j^{-1} U(\mathbf{X}_i)^T U(\mathbf{X}_j).$$

By Condition (C1) and (C3),  $E(r_i^{-2}) = E(||\boldsymbol{\varepsilon}_i||^{-2})(1 + o(1))$ ,  $\boldsymbol{\varepsilon}_i = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})$ . So the power function of  $T_n$  is

$$\beta_{\text{os}}(||\boldsymbol{\theta}||) = \Phi \left( -z_\alpha + E(||\boldsymbol{\varepsilon}||^{-2}) \frac{pn \boldsymbol{\theta}^T \boldsymbol{\theta}}{\sqrt{2\text{tr}(\boldsymbol{\Sigma}^2)}} \right).$$

Chen and Qin (2010) and Wang, et al. (2015) show that the asymptotic power of their proposed tests are

$$\begin{aligned}\beta_{\text{CQ}}(||\boldsymbol{\theta}||) &= \Phi \left( -z_\alpha + \frac{np\boldsymbol{\theta}^T\boldsymbol{\theta}}{E(||\boldsymbol{\varepsilon}||^2)\sqrt{2\text{tr}(\boldsymbol{\Sigma}^2)}} \right), \\ \beta_{\text{SS}}(||\boldsymbol{\theta}||) &= \Phi \left( -z_\alpha + (E(||\boldsymbol{\varepsilon}||^{-1}))^2 \frac{np\boldsymbol{\theta}^T\boldsymbol{\theta}}{\sqrt{2\text{tr}(\boldsymbol{\Sigma}^2)}} \right).\end{aligned}$$

Thus, the asymptotic relative efficiency of our proposed test with respect to these two tests are

$$\begin{aligned}\text{ARE}(\text{OS}, \text{CQ}) &= E(||\boldsymbol{\varepsilon}||^{-2})E(||\boldsymbol{\varepsilon}||^2) \geq 1 \\ \text{ARE}(\text{OS}, \text{SS}) &= \frac{E(||\boldsymbol{\varepsilon}||^{-2})}{\{E(||\boldsymbol{\varepsilon}||^{-1})\}^2} = 1 + \frac{\text{var}(||\boldsymbol{\varepsilon}||^{-1})}{\{E(||\boldsymbol{\varepsilon}||^{-1})\}^2} \geq 1.\end{aligned}$$

Both of the above two equations only hold when  $||\boldsymbol{\varepsilon}||/E(||\boldsymbol{\varepsilon}||) \xrightarrow{p} 1$ . If  $||\boldsymbol{\varepsilon}||/E(||\boldsymbol{\varepsilon}||) \xrightarrow{p} 1$ , these three tests are asymptotic equivalent. Otherwise, our proposed test would perform better than the other two tests.

When  $\boldsymbol{\varepsilon}_i \sim N(0, \mathbf{I}_p)$ ,  $||\boldsymbol{\varepsilon}_i||/\sqrt{p} \xrightarrow{p} 1$ . Then,  $\text{ARE}(\text{OS}, \text{CQ})$  and  $\text{ARE}(\text{OS}, \text{SS})$  are all equal to one.

When  $\boldsymbol{\varepsilon}_i \sim t_p(0, \mathbf{I}_p, v)$ , where  $t_p(0, \mathbf{I}_p, v)$  is the standard  $p$ -dimensional multivariate  $t$  distribution with  $v$  degrees of freedom, we have

$$\text{ARE}(\text{OS}, \text{CQ}) = \frac{v}{v-2}, \quad \text{ARE}(\text{OS}, \text{SS}) = \frac{v\Gamma^2(v/2)}{2\Gamma^2((v+1)/2)}.$$

In this case,  $\psi_g(t) = (p+v)t/(v+t^2) \rightarrow pt^{-1}$  as  $t \rightarrow \infty$ . So, our uniformly optimal weighted function  $K(t)$  would be consistent with the “optimal” weighted function  $\psi_g(t)$ .

When  $\boldsymbol{\varepsilon}_i$  is from the mixtures of two multivariate normal distributions  $MN(\kappa, \sigma, \mathbf{I}_p)$  with density function  $(1-\kappa)f_p(0, \mathbf{I}_p) + \kappa f_p(0, \sigma^2\mathbf{I}_p)$ , where  $f_p(\cdot)$  is the density function of  $p$ -variate multivariate normal distribution, we have

$$\text{ARE}(\text{OS}, \text{CQ}) = (1-\kappa + \kappa/\sigma^2)(1-\kappa + \kappa\sigma^2), \quad \text{ARE}(\text{OS}, \text{SS}) = \frac{1-\kappa + \kappa/\sigma^2}{(1-\kappa + \kappa/\sigma)^2}.$$

As  $\sigma^2 \rightarrow \infty$ ,  $\text{ARE}(\text{OS}, \text{CQ})$  will be arbitrary large and  $\text{ARE}(\text{OS}, \text{SS})$  will converge to  $1/(1-\kappa)$ . However, in this case,  $\psi_g(t) = \frac{(1-\kappa)t \exp(-t^2/2) + \sigma^{-3}\kappa t \exp(-t^2/(2\sigma^2))}{(1-\kappa) \exp(-t^2/2) + \sigma^{-3}\kappa \exp(-t^2/(2\sigma^2))} \rightarrow t$  as  $t \rightarrow \infty$ , which

is consistent with Chen and Qin (2010)'s test. So,  $K(t) = \psi_g(t)$  would not be optimal in such case. Thus, for high dimensional data, a simply extension of  $Q_n$  with  $\psi_g(t)$  may not be always the best test.

Table 1 reports asymptotic relative efficiency between these three tests under the multivariate  $t$ -distributions with different degrees of freedom and mixture normal distributions. Formulas of asymptotic relative efficiency with these two distributions are given in the Supplementary Material.

Table 1: Asymptotic relative efficiencies with different distributions.

	$t_p(0, \mathbf{I}_p, 3)$	$t_p(0, \mathbf{I}_p, 4)$	$t_p(0, \mathbf{I}_p, 5)$	$t_p(0, \mathbf{I}_p, 6)$	$N(0, \mathbf{I}_p)$	$MN(0.2, 3, \mathbf{I}_p)$	$MN(0.2, 10, \mathbf{I}_p)$	$MN(0.8, 10, \mathbf{I}_p)$
ARE(SS,CQ)	2.54	1.76	1.51	1.38	1.00	2.06	13.98	6.28
ARE(OS,CQ)	3.00	2.00	1.67	1.50	1.00	2.25	16.68	16.68
ARE(OS,SS)	1.18	1.13	1.11	1.09	1.00	1.09	1.19	2.65

$t_p(0, \Lambda, v)$ ,  $p$ -dimensional multivariate  $t$  distribution with  $v$  degrees of freedom and scatter matrix  $\Lambda$ ;  $MN(\kappa, \sigma, \Lambda)$ , mixture multivariate normal distribution with density function  $(1 - \kappa)f_p(0, \Lambda) + \kappa f_p(0, \sigma^2 \Lambda)$ , where  $f_p(\cdot)$  is the density function of  $p$ -variate multivariate normal distribution.

### 3 Simulation

Here we report a simulation study designed to evaluate the performance of the proposed test. All the simulation results are based on 2,500 replications. We consider the following five elliptical distributions: (I)  $N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ ; (II)  $t_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, 3)$ ; (III)  $t_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, 4)$ ; (IV)  $MN(0.2, 10, \boldsymbol{\Sigma})$ ; (V)  $MN(0.8, 10, \boldsymbol{\Sigma})$  and two independent component model  $\mathbf{X}_i = \boldsymbol{\Sigma}^{1/2} \mathbf{Z}_i + \boldsymbol{\mu}$ ,  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})$  where (VI)  $Z_{ij} \sim t_3$ ; (VII)  $Z_{ij} \sim 0.8N(0, 1) + 0.2N(0, 100)$ . The scatter matrix is  $\boldsymbol{\Sigma} = (0.5^{|i-j|})$ . The sample size is  $n = 40$  and the dimension is  $p = 200, 400, 800$ . Under the alternative hypothesis, two patterns of allocation are considered: (Dense case): the first 50% components of  $\boldsymbol{\theta}$  are zeros; (Sparse case) the first 95% components of  $\boldsymbol{\theta}$  are

zeros. And we fixed  $\boldsymbol{\theta}^T \boldsymbol{\theta} / \sqrt{\text{tr}(\boldsymbol{\Sigma})} = 0.1$  for the first four scenarios (I)-(IV) and (VI), and  $\boldsymbol{\theta}^T \boldsymbol{\theta} / \sqrt{\text{tr}(\boldsymbol{\Sigma})} = 1$  for scenario (V) and (VII). We compare our proposed test with Chen and Qin (2010)'s test and Wang, et al. (2015)'s test. Table 2 reports the empirical sizes and power of these three tests. All these tests can control the empirical sizes very well. For multivariate normal distribution and independent component model, the difference between these three tests are negligible. It is not strange because  $\|\boldsymbol{\varepsilon}\|/\sqrt{p} \xrightarrow{p} 1$  in this case. Then, the asymptotic relative efficiency between these tests are all one. But under the non-normal cases, both Wang, et al. (2015)'s test and our proposed test performs better than Chen and Qin (2010)'s test in all cases. For heavy-tailed distributions, those direction-based tests will perform better than those moment-based tests. Furthermore, our proposed test is more powerful than Wang, et al. (2015)'s test in these cases, which is consistent with the asymptotic analysis. Though Wang, et al. (2015)'s test is very powerful method, it loses all the information of the module of the observations. All these results suggest that our proposed test is very efficient and robust in a wide range of distributions.

## 4 Discussion

In this paper, we propose a weighted sign test and determine the “optimal” weight function by maximizing the power function. Our asymptotic and numerical results together suggest that the proposed optimal sign test is quite robust and efficient in testing the population mean vector. This article concerns the one sample location problem. Testing the equality of two sample locations are also a very important problem (Srivastava and Du, 2008; Cai, Liu and Xia, 2014; Chen et al., 2011; Gregory et al., 2015). In the two sample problem, the common mean vector is not specified and need to be estimated. How to extend our method deserves further study. Furthermore, the proposed test procedure is essentially developed under the framework of  $L_2$ -norm-based tests. In another direction, Cai, Liu and Xia (2014) and Zhong, Chen and Xu (2013) used the max-norm or thresholding approach to construct tests rather than the  $L_2$ -norm. Generally speaking, the max-norm test is for



Table 2: Empirical sizes and power (%) comparison at 5% significance under Scenarios (I)-(V)

	Size			Dense			Sparse		
	CQ	SS	OS	CQ	SS	OS	CQ	SS	OS
$(n, p) = (40, 200)$									
(I)	5.8	6.3	6.2	74.9	76.6	76.0	81.0	83.5	82.8
(II)	4.5	5.7	6.2	32.4	68.2	75.3	33.6	72.9	78.7
(III)	5.1	5.9	5.7	43.1	68.9	75.2	46.3	77.4	82.3
(IV)	6.1	7.1	6.2	9.0	55.1	63.7	10.3	60.6	68.9
(V)	6.1	7.0	5.4	12.6	58.6	94.7	13.4	64.1	96.3
(VI)	6.6	7.3	5.4	25.1	29.7	29.5	27.4	34.0	34.3
(VII)	4.8	5.1	4.8	34.8	38.6	39.4	40.9	45.3	45.1
$(n, p) = (40, 400)$									
(I)	5.2	6.0	5.9	78.6	80.1	79.9	80.3	82.6	82.3
(II)	4.3	5.1	4.7	29.7	68.1	76.9	31.9	70.7	79.4
(III)	4.9	6.0	6.6	40.8	73.7	80.5	43.1	76.6	80.9
(IV)	5.4	6.5	5.3	8.3	54.5	65.3	8.5	59.0	68.3
(V)	4.7	6.9	5.1	10.6	57.9	95.2	10.6	59.9	94.6
(VI)	3.2	4.5	4.7	23.3	27.2	27.4	24.2	27.0	26.4
(VII)	6.0	7.0	5.8	34.8	39.9	39.7	38.4	41.4	41.9
$(n, p) = (40, 800)$									
(I)	4.2	5.8	5.4	80.7	82.4	81.5	78.4	80.5	80.1
(II)	5.3	5.1	5.4	31.7	69.1	77.5	31.3	72.1	79.7
(III)	5.2	5.2	5.7	43.9	74.3	80.2	44.5	74.2	81.7
(IV)	4.1	4.7	5.5	6.4	54.2	65.7	7.3	57.8	68.1
(V)	5.9	7.0	5.0	10.3	59.9	94.8	9.6	60.2	94.7
(VI)	4.3	5.1	5.3	21.3	25.5	26.4	21.7	25.8	26.7
(VII)	4.7	5.7	5.4	36.8	41.0	40.1	36.3	40.6	40.7

CQ, Chen and Qin (2010)'s test; SS, Wang, et al. (2015)'s test; OS, our proposed high dimensional uniformly optimal sign test.

more sparse and stronger signals whereas the  $L_2$ -norm test is for denser but fainter signals. Fan, Liao and Yao (2015) also proposed a power-enhancement test based on a screening technique. Developing a spatial-sign-based test for sparse signals is of interest in the future study.

## Appendix A: Scalar-invariant test

Here we replace  $\Sigma$  in  $R_n$  with its diagonal matrix and define the following test statistic

$$T_n = \frac{2}{n(n-1)} \sum_{i < j} K(\|\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_i\|) K(\|\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_j\|) U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_i)^T U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_j),$$

where  $\hat{\mathbf{D}}_{ij}$  is the corresponding diagonal matrix estimator using leave-two-out sample  $\{\mathbf{X}_k\}_{k \neq i,j}$  in Feng, et al. (2015a). Now,  $T_n$  is invariant under scalar transformations  $\mathbf{X}_i \rightarrow \mathbf{B}\mathbf{X}_i$ ,  $\mathbf{B} = \text{diag}\{b_1^2, \dots, b_p^2\}$ . Define  $\mathbf{R} = \mathbf{D}^{-1/2} \Sigma \mathbf{D}^{-1/2}$  where  $\mathbf{D}$  is the diagonal matrix of  $\Sigma$ . Now the conditions (C1)-(C3) become

$$(C1') \quad \text{tr}(\mathbf{R}^4) = o(\text{tr}^2(\mathbf{R}^2)) \text{ and } \text{tr}(\mathbf{R}^2) - p = o(n^{-1}p^2).$$

$$(C2') \quad \tilde{\nu}_4 = O(\tilde{\nu}_2^2) \text{ where } \tilde{\nu}_l = E(K^l(\tilde{r}_i)) \text{ and } \tilde{r}_i = \|\mathbf{D}^{-1/2} \mathbf{X}_i\|.$$

$$(C3') \quad \boldsymbol{\theta}^T \mathbf{D}^{-1} \boldsymbol{\theta} = O(\tilde{c}_0^{-2} \tilde{\sigma}_n), \quad \boldsymbol{\theta}^T \mathbf{D}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2} \boldsymbol{\theta} = o(np \tilde{c}_0^{-2} \tilde{\sigma}_n) \text{ where } \tilde{c}_0 = E\{K(\tilde{r}_i) \tilde{r}_i^{-1}\} \text{ and } \tilde{\sigma}_n^2 = 2n^{-2}p^{-2} \tilde{\nu}_2^2 \text{tr}(\mathbf{R}^2).$$

Furthermore, we need another technical condition for the consistency of  $\hat{\mathbf{D}}_{ij}$ .

$$(C4') \quad n^{-2}p^2/\text{tr}(\mathbf{R}^2) = O(1) \text{ and } \log(p) = o(n).$$

**Theorem 3** *Under Conditions (C1')-(C4'), as  $n, p \rightarrow \infty$ , we have*

$$\frac{T_n - \tilde{c}_0^2 \boldsymbol{\theta}^T \mathbf{D}^{-1} \boldsymbol{\theta}}{\tilde{\sigma}_n} \xrightarrow{d} N(0, 1).$$

Correspondingly, the ratio-consistent estimator of  $\tilde{\sigma}_n^2$  is

$$\begin{aligned} \tilde{\sigma}_n^2 = & 2n^{-4} \sum_{i \neq j} \sum K^2(\|\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_i\|) K^2(\|\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_j\|) \{U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_i) - \tilde{\boldsymbol{\mu}}_{i,j}\}^T U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_j) \\ & \times \{U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_j) - \tilde{\boldsymbol{\mu}}_{i,j}\}^T U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_i), \end{aligned}$$

where  $\tilde{\boldsymbol{\mu}}_{i,j} = \frac{1}{n-2} \sum_{k \neq i,j} U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_k)$ .

So the asymptotic power function of  $T_n$  is

$$\beta_{T_n}(\|\boldsymbol{\theta}\|) = \Phi \left( -z_\alpha + \frac{[E\{K(\tilde{r}_i)\tilde{r}_i^{-1}\}]^2 pn \boldsymbol{\theta}^T \mathbf{D}^{-1} \boldsymbol{\theta}}{E\{K^2(\tilde{r}_i)\} \sqrt{2\text{tr}(\mathbf{R}^2)}} \right).$$

By the Cauchy inequality, the optimal weighted function is also  $K(t) = t^{-1}$ .

## Appendix B: Technical Details

Define  $\mathbf{U}_i = U(\mathbf{X}_i - \boldsymbol{\theta})$ ,  $\mathbf{u}_i = U(\boldsymbol{\varepsilon}_i)$ ,  $r_i^* = \|\boldsymbol{\varepsilon}_i\|$ . First, we restate Lemma 4 in Zou et al. (2014).

**Lemma 1** *Suppose  $\mathbf{u}$  are independent identically distributed uniform on the unit  $p$  sphere. For any  $p \times p$  symmetric matrix  $\mathbf{M}$ , we have*

$$\begin{aligned} E(\mathbf{u}^T \mathbf{M} \mathbf{u})^2 &= \{\text{tr}^2(\mathbf{M}) + 2\text{tr}(\mathbf{M}^2)\} / (p^2 + 2p), \\ E(\mathbf{u}^T \mathbf{M} \mathbf{u})^4 &= \{3\text{tr}^2(\mathbf{M}^2) + 6\text{tr}(\mathbf{M}^4)\} / \{p(p+2)(p+4)(p+6)\}. \end{aligned}$$

### B1: Proof of Theorem 1

Obviously,  $E(W_n) = 0$  and

$$\text{var}(W_n) = \frac{2}{n(n-1)} E\{K^2(r_i) K^2(r_j) (\mathbf{U}_i^T \mathbf{U}_j)^2\}$$

Because  $\|\mathbf{X}_i\|^2 = \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i^T (\boldsymbol{\Sigma} - \mathbf{I}_p) \boldsymbol{\varepsilon}_i$  and  $E\{\boldsymbol{\varepsilon}_i^T (\boldsymbol{\Sigma} - \mathbf{I}_p) \boldsymbol{\varepsilon}_i\} = E(\|\boldsymbol{\varepsilon}_i\|^2) p^{-1} \{\text{tr}(\boldsymbol{\Sigma}^2) - p\}$ , So  $\|\mathbf{X}_i\| = \|\boldsymbol{\varepsilon}_i\|(1 + o_p(1))$ . Similarly,  $\mathbf{U}_i = \boldsymbol{\Sigma}^{1/2} \mathbf{u}_i(1 + o_p(1))$ . Thus,

$$\begin{aligned} \text{var}(W_n) &= 2n^{-2} E\{K^2(r_i^*) K^2(r_j^*) (\mathbf{u}_i^T \boldsymbol{\Sigma} \mathbf{u}_j)^2\} (1 + o(1)) \\ &= 2n^{-2} p^{-2} \nu_2^2 \text{tr}(\boldsymbol{\Sigma}^2) (1 + o(1)). \end{aligned}$$

Thus, we only need to proof the normality of  $W_n$ . Define  $W_{nk} = \sum_{i=2}^k Z_{ni}$  where  $Z_{ni} = \sum_{j=1}^{i-1} \frac{1}{n(n-1)} \mathbf{V}_i^T \mathbf{V}_j$ ,  $\mathbf{V}_i = K(r_i) \mathbf{U}_i$ . Let  $\mathbf{A} = E(\mathbf{V}_i \mathbf{V}_i^T)$ . Let  $\mathcal{F}_{n,i} = \sigma\{\mathbf{V}_1, \dots, \mathbf{V}_i\}$  be the  $\sigma$ -field generated by  $\{\mathbf{V}_j, j \leq i\}$ . Obviously,  $E(Z_{ni} | \mathcal{F}_{n,i-1}) = 0$  and it follows that  $\{W_{nk}, \mathcal{F}_{n,k}; 2 \leq k \leq n\}$  is a zero mean martingale. The central limit theorem (Hall and Hyde, 1980) will hold if we can show

$$\frac{\sum_{j=2}^n E(Z_{nj}^2 | \mathcal{F}_{n,j-1})}{\sigma_n^2} \xrightarrow{p} 1. \quad (2)$$

and for any  $\epsilon > 0$ ,

$$\sigma_n^{-2} \sum_{j=2}^n E\{Z_{nj}^2 I(|Z_{nj}| > \epsilon \sigma_n) | \mathcal{F}_{n,j-1}\} \xrightarrow{p} 0. \quad (3)$$

It can be shown that

$$\begin{aligned} \sum_{j=2}^n E(Z_{nj}^2 | \mathcal{F}_{n,j-1}) &= \frac{4}{n^2(n-1)^2} \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbf{V}_i^T \mathbf{A} \mathbf{V}_i \\ &\quad + \frac{4}{n^2(n-1)^2} \sum_{j=2}^n \sum_{i_1 < i_2}^{j-1} \mathbf{V}_{i_1}^T \mathbf{A} \mathbf{V}_{i_2} \\ &\doteq C_{n1} + C_{n2} \end{aligned}$$

Obviously,  $E(C_{n1}) = \frac{2}{n(n-1)} \text{tr}(\mathbf{A}^2) = \sigma_n^2(1 + o(1))$  by the calculation of  $\text{var}(W_n)$ . And  $\text{var}(C_{n1}) = O(n^{-5}) \text{var}((\mathbf{V}_i^T \mathbf{A} \mathbf{V}_i)^2)$ . According to Lemma 1, we have  $\text{var}((\mathbf{V}_i^T \mathbf{A} \mathbf{V}_i)^2) = O(\text{tr}^2(\mathbf{A}^2) + \text{tr}(\mathbf{A}^4))$ . Thus, by Condition (C1), we have  $\text{var}(C_{n1}) = O(n^{-5}) \text{tr}^2(\mathbf{A}^2) = o(\sigma_n^4)$ . Thus,  $C_{n1}/\sigma_n^2 \xrightarrow{p} 1$ . Similarly,  $E(C_{n2}) = O(n^{-4}) \text{tr}(\mathbf{A}^4) = o(\sigma_n^4)$ . Then (2) holds. Next, to proof (3), by Chebyshev's inequality, we only need to show

$$E \left\{ \sum_{j=2}^n E(Z_{nj}^4 | \mathcal{F}_{n,j-1}) \right\} = o(\sigma_n^4).$$

Note that

$$E \left\{ \sum_{j=2}^n E(Z_{nj}^4 | \mathcal{F}_{n,j-1}) \right\} = \sum_{j=2}^n E(Z_{nj}^4) = O(n^{-8}) \sum_{j=2}^n E \left( \sum_{i=1}^{j-1} \mathbf{V}_j^T \mathbf{V}_i \right)^4.$$

which can be decomposed as  $3Q + P$  where

$$Q = O(n^{-8}) \sum_{j=2}^n \sum_{s < t}^{j-1} E(\mathbf{V}_j^T \mathbf{V}_s \mathbf{V}_s^T \mathbf{V}_j \mathbf{V}_j^T \mathbf{V}_t \mathbf{V}_t^T \mathbf{V}_j)$$

$$P = O(n^{-8}) \sum_{j=2}^n \sum_{i=1}^{j-1} E\{(\mathbf{V}_j^T \mathbf{V}_i)^4\}$$

Obviously,  $Q = O(n^{-5})E((\mathbf{V}_j^T \mathbf{A} \mathbf{V}_j)^2) = O(n^{-5})\text{tr}^2(\mathbf{A}^2)$  by Lemma 1 and Condition (C1). Then  $Q = o(\sigma_n^4)$ . Similarly, we can show that  $P = O(n^{-6})\text{tr}^2(\mathbf{A}^2) = o(\sigma_n^4)$ . Here we complete the proof.  $\square$

## B2: Proof of Theorem 2

By the Taylor expansion, we have

$$U(\mathbf{X}_i) = \mathbf{U}_i + r_i^{-1}(\mathbf{I}_p - \mathbf{U}_i \mathbf{U}_i^T) \boldsymbol{\theta} + o_p(n^{-1}).$$

Thus, taking the same procedure as Theorem 1, we have

$$W_n = \frac{2}{n(n-1)} \sum_{i < j} \mathbf{V}_i^T \mathbf{V}_j + \frac{2}{n(n-1)} \sum_{i < j} K(r_i) r_i^{-1} \mathbf{V}_j^T \boldsymbol{\theta}$$

$$+ \frac{2}{n(n-1)} \sum_{i < j} r_i^{-1} r_j^{-1} K(r_i) K(r_j) \boldsymbol{\theta}^T \boldsymbol{\theta} + o_p(\sigma_n)$$

And

$$E \left( \frac{2}{n(n-1)} \sum_{i < j} K(r_i) r_i^{-1} \mathbf{V}_j^T \boldsymbol{\theta} \right)^2 = O(n^{-2} p^{-1} c_0^2 \boldsymbol{\theta}^T \boldsymbol{\Sigma} \boldsymbol{\theta}) = o(\sigma_n^2)$$

by Condition (C3). Similarly,

$$\frac{2}{n(n-1)} \sum_{i < j} r_i^{-1} r_j^{-1} K(r_i) K(r_j) \boldsymbol{\theta}^T \boldsymbol{\theta} = c_0^2 \boldsymbol{\theta}^T \boldsymbol{\theta} + o_p(\sigma_n).$$

Then,

$$W_n = \frac{2}{n(n-1)} \sum_{i < j} \mathbf{V}_i^T \mathbf{V}_j + c_0^2 \boldsymbol{\theta}^T \boldsymbol{\theta} + o_p(\sigma_n).$$

According to Theorem 1, we can easily obtain the result.  $\square$

### B3: Consistency of $\hat{\sigma}_n^2$

Taking the same procedure as the proof of Theorem 2 in Chen and Qin (2010), we have

$$\begin{aligned}\hat{\sigma}_n^2 &= 2n^{-4} \sum_{i \neq j} K^2(r_i) K^2(r_j) (\mathbf{U}_i^T \mathbf{U}_j)^2 + o_p(\sigma_n^2) \\ &= 2n^{-4} \sum_{i \neq j} (\mathbf{V}_i^T \mathbf{V}_j)^2 + o_p(\sigma_n^2),\end{aligned}$$

by Condition (C3). According to the proof of Theorem 1, we have  $E((\mathbf{V}_i^T \mathbf{V}_j)^2) = \text{tr}(\mathbf{A}^2) = p^{-2} \nu_2^2 \text{tr}(\mathbf{\Sigma}^2)(1+o(1))$ . So  $E(\hat{\sigma}_n^2) = \sigma_n^2(1+o(1))$ . And  $\text{var}((\mathbf{V}_i^T \mathbf{V}_j)^2) = o(\text{tr}^2 \mathbf{A}^2)$  by Condition (C1) and (C2). Thus,  $\text{var}(\hat{\sigma}_n^2) = o(\sigma_n^4)$ . So  $\hat{\sigma}_n^2 / \sigma_n^2 \xrightarrow{p} 1$ .  $\square$

### B4: Proof of Theorem 3

By the Tyler's expansion,

$$\begin{aligned}U(\hat{\mathbf{D}}_{ij}^{-1/2} \mathbf{X}_i) &= \mathbf{U}_i - (\mathbf{I}_p - \mathbf{U}_i \mathbf{U}_i^T)(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}^{-1/2}) \mathbf{U}_i \\ &\quad + \tilde{r}_i^{-1} (\mathbf{I}_p - \mathbf{U}_i \mathbf{U}_i^T) \mathbf{D}^{-1/2} \boldsymbol{\theta} + o_p(n^{-1}).\end{aligned}$$

Taking the same procedure as the proof of Theorem 1 in Feng and Sun (2015), by Conditions (C1'), (C2') and (C4'), we have

$$\begin{aligned}T_n &= \frac{2}{n(n-1)} \sum_{i < j} K(\tilde{r}_i) K(\tilde{r}_j) \mathbf{u}_i^T \mathbf{\Sigma}^{1/2} \mathbf{D}^{-1} \mathbf{\Sigma}^{1/2} \mathbf{u}_j \\ &\quad + \frac{2}{n(n-1)} \sum_{i < j} K(\tilde{r}_i) \tilde{r}_i^{-1} \mathbf{U}_j^T (\mathbf{I}_p - \mathbf{U}_i \mathbf{U}_i^T) \mathbf{D}^{-1/2} \boldsymbol{\theta} \\ &\quad + \frac{2}{n(n-1)} \sum_{i < j} K(\tilde{r}_j) \tilde{r}_j^{-1} \mathbf{U}_i^T (\mathbf{I}_p - \mathbf{U}_j \mathbf{U}_j^T) \mathbf{D}^{-1/2} \boldsymbol{\theta} \\ &\quad + \frac{2}{n(n-1)} \sum_{i < j} K(\tilde{r}_i) K(\tilde{r}_j) \tilde{r}_i^{-1} \tilde{r}_j^{-1} \boldsymbol{\theta}^T \mathbf{D}^{-1/2} (\mathbf{I}_p - \mathbf{U}_i \mathbf{U}_i^T) (\mathbf{I}_p - \mathbf{U}_j \mathbf{U}_j^T) \mathbf{D}^{-1/2} \boldsymbol{\theta} + o_p(n^{-2}) \\ &\doteq T_{n1} + T_{n2} + T_{n3} + T_{n4}.\end{aligned}$$

By the same arguments as the proof of Theorem 1, we have

$$T_{n1} / \tilde{\sigma}_n \xrightarrow{d} N(0, 1).$$

and

$$E(T_{n2}^2) = E(T_{n3}^2) = O(n^{-1}p^{-1}\tilde{c}_0^2\boldsymbol{\theta}^T\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}\boldsymbol{\theta}), \quad T_{n4} = \tilde{c}_0^2\boldsymbol{\theta}^T\mathbf{D}^{-1}\boldsymbol{\theta} + o_p(\tilde{\sigma}_n).$$

Thus, by Condition (C3'),  $(T_n - \tilde{c}_0^2\boldsymbol{\theta}^T\mathbf{D}^{-1}\boldsymbol{\theta})/\tilde{\sigma}_n \xrightarrow{d} N(0, 1)$ . □

## References

- BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica* **6**, 311–29.
- Cai, T. T., Liu, W. and Xia, Y. (2014), Two-Sample Test of High Dimensional Means under Dependence, *J. Roy. Statist. Soc. Ser. B*, 76, 349–372.
- Chen, L. S., Paul, D., Prentice, R. L. and Wang, P. (2011), A Regularized Hotelling's  $T^2$  Test for Pathway Analysis in Proteomic Studies, *J. Am. Statist. Assoc.*, 106, 1345–1360.
- CHEN, S. X. and QIN, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**, 808–835.
- Fan, J., Liao, Y., and Yao, J. (2015). Power Enhancement in High Dimensional Cross-Sectional Tests, *arXiv:1310.3899*.
- FENG, L. AND SUN, F. S. (2015). Spital-sign based high dimensional location test. *Manuscript*.
- FENG, L., ZOU, C. AND WANG, Z. (2015). Multivariate-sign-based high-dimensional tests for the two-sample location problem, *J. Amer. Statist. Assoc.*, To appear.
- FENG, L., ZOU, C., WANG, Z. AND ZHU, L. X. (2014). Two sample Behrens-Fisher problem for high-dimensional data, *Statist. Sinica*, To appear.
- Gregory, K. B., Carroll, R. J., Baladandayuthapani, V. and Lahiri, S. N. (2015), A Two-Sample Test For Equality of Means in High Dimension, , *J. Am. Statist. Assoc.*, To appear.

- Hall, P. G. and Hyde, C. C. (1980), *Martingale Central Limit Theory and its Applications*, New York: Academic Press.
- HALLIN, M. AND PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-mahalanobis ranks, *Ann. Statist.*, **30**, 1103–1133.
- Hallin, M. and Paindaveine, D. (2006), Semiparametrically Efficient Rank-based Inference for Shape. I: Optimal Rank-Based Tests for Sphericity, *Ann. Statist.*, **34**, 2707–2756.
- LE CAM, L. (1986). *Asymptotic methods in statistical decision theory*. Springer, New York.
- OJA, H. (2010). *Multivariate Nonparametric Methods with R*. Springer, New York.
- Paindaveine, D. and Verdebout, T. (2015), On High-Dimensional Sign Tests, *Bernoulli*, To appear.
- PARK, J. AND AYYALA, D. N. (2013). A test for the mean vector in large dimension and small samples, *J. Stat. Plan. Infer.*, **143**, 929–943.
- PURI, M. L. AND SEN, P. K. (1971). *Nonparametric methods in multivariate analysis*, New York: Wiley.
- RANGLES, R. H. (1992). A two sample extension of the multivariate interdirection sign test, *L1-Statistical Analysis and Related Methods* (ed. Y. Dodge), Elsevier, Amsterdam, pp. 295–302.
- SRIVASTAVA, M.S. (2009). A test of the mean vector with fewer observations than the dimension under non-normality, *J. Multivar. Anal.*, **100**, 518–532.
- Srivastava, M. S. and Du, M. (2008), A Test for the Mean Vector with Fewer Observations than the Dimension, *J. Multivar. Anal.*, **99**, 386–402.
- Srivastava, M. S., Katayama, S. and Kano, Y. (2013), A Two Sample Test in High Dimensional Data, *J. Multivar. Anal.*, **114**, 349–358.



- WANG, L., PENG, B. AND LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector, *J. Amer. Statist. Assoc.*, To appear.
- Zhong, P., Chen, S. X. and Xu, M. (2013), Tests Alternative to Higher Criticism for High Dimensional Means under Sparsity and Column-Wise Dependence, *Ann. Statist.*, 41, 2703–3110.
- ZOU, C., PENG, L, FENG, L. AND WANG, Z. (2014). Multivariate-sign-based high-dimensional tests for sphericity, *Biometrika*, **101**, 229–236.